

Utah State University

DigitalCommons@USU

All Graduate Plan B and other Reports

Graduate Studies

5-2014

Predicting Tourist Inflows to Punta Cana, Dominican Republic, Using Google Trends

Michelle E. De La Oz Pineda
Utah State University

Follow this and additional works at: <https://digitalcommons.usu.edu/gradreports>



Part of the [Economics Commons](#)

Recommended Citation

De La Oz Pineda, Michelle E., "Predicting Tourist Inflows to Punta Cana, Dominican Republic, Using Google Trends" (2014). *All Graduate Plan B and other Reports*. 360.

<https://digitalcommons.usu.edu/gradreports/360>

This Report is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Plan B and other Reports by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.



PREDICTING TOURIST INFLOWS TO PUNTA CANA, DOMINICAN
REPUBLIC, USING GOOGLE TRENDS

by

Michelle E. De La Oz Pineda

A Plan B Report submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Economics

Approved:

Devon Gorry
Major Professor

William F. Shughart II
Committee Member

Diana Thomas
Committee Member

UTAH STATE UNIVERSITY
Logan, Utah

2014

ABSTRACT

Predicting Tourist Inflows to Punta Cana, Dominican Republic using Google Trends

by

Michelle E. De La Oz, Master of Science

Utah State University, 2014

Major Professor: Dr. Devon Gorry
Department: Economics and Finance

This paper seeks to demonstrate that the incorporation of search statistics in autoregressive models used to predict the arrivals of tourists to Punta Cana, Dominican Republic, improve the predictive power of the models. This paper explores whether the Internet search information in “Punta Cana” in the United States and Canada between January 2004 and August 2013, reflects the behavior of this variable in real time by using nowcasting methodology that combines variables with different frequencies of time. We find that including the searches of “Punta Cana” that Canadians make on Google in the first week of a month helps predict the actual arrival of Canadian tourists to Punta Cana in that month. The same applies in the case of Americans, but using the searches made during the third week of a month.

ACKNOWLEDGMENTS

I would like to thank my Major Thesis Professor, Dr. Devon Gorry, for providing the continuous encouragement, support and advise for me to make this thesis possible. Special thanks for my committee, Dr. William F. Shughart II and Dr. Diana Thomas, for their guidance and helpful suggestions. I thank Faruk Miguel for the invaluable support and for always believing in me.

I give heartfelt thanks to my family and friends for their encouragement and moral support during the entire thesis process. I could not have done it without all of you.

Michelle E. De La Oz Pineda

CONTENTS

i.	ABSTRACT	ii
ii.	ACKNOWLEDGEMENTS.....	iii
I.	INTRODUCTION	1
II.	LITERATURE REVIEW.....	4
III.	METHODOLOGY	9
	3.1 Data	9
	3.2 The Google Index.....	12
	3.3 Internet Usage.....	14
IV.	MODEL SELECTION	16
V.	ESTIMATION OF THE MODELS.....	20
VI.	CONCLUSIONS	23
VII.	REFERENCES.....	25
VIII.	APPENDIX.....	28

I. INTRODUCTION

It has been shown that the use of trend information from Internet search activity improves the accuracy of immediate forecasting. This paper examines the relationship between Internet search activity and the arrival of tourists to Punta Cana, Dominican Republic, between January 2004 and August 2013. The objective is to examine whether Internet visits are related to the actual visits and demonstrate that Google searches are an effective predictor of the arrival of tourists to Punta Cana.

The Dominican Republic is a middle-income developing country primarily dependent on trade and services, especially tourism. Tourism is one of the most important economic activities of the Dominican Republic, representing 6.7% of GDP in the studied period. Therefore, predicting tourist arrivals is of major relevance to businesses and policy-makers in this country. The paper seeks to determine the extent to which Internet search of tourism related queries of Punta Cana can improve forecasts of tourist arrivals, in order to enable policy-makers and entrepreneurs related to the tourism sector to make more informed decisions on tourism in the Dominican Republic.

Google has developed a tool called Google Trends that helps identify key words with the most search volume. This information reveals the interests of Google users, but also their intentions and future actions. Using data from search queries on Google, in Canada and the United States, we test whether including this information in our model improves the efficiency adjustment of nowcasting models for the arrival of tourists to Punta Cana, Dominican Republic.

Nowcasting is relevant in economics because key statistics on the present state of the economy are available only with a significant delay. The research interest is to show that web search queries improve the predictive power of our model, with the goal that policy makers are able to make strategic decisions having access to the information they need more quickly. The aim of this paper is to see whether that the use of information technology has played a new role in the improvement of the estimates made by these prediction models.

We find that tourism in Punta Cana can be accurately predicted when including Google Trends data in the prediction models. Additionally, we found that we can get timelier information than the official monthly report of the Central Bank of the Dominican Republic. For example, the volume of searches in Canada of *Punta Cana* during the first week of September is helpful in predicting Canadian tourist arrivals to Punta Cana that same month, figures that are released several weeks later in October. We find that queries can be useful leading indicators for touristic arrivals to Punta Cana, based on consumer's traveling behavior, where tourists start

planning their travels significantly in advance of their actual travel date. The same conclusions apply for the US case, but by using the search queries during the third week of the month.

The paper is composed of five parts. First we present the existing economic literature that includes search queries in their studies. Then we lay out the methodology we will use to answer the question posed in this research, followed by an explanation of the origin and characteristics of the data we used. Third, we present the econometric model to be estimated. The fourth section presents the estimated model and the fifth section provides the conclusions.

II. LITERATURE REVIEW

This research follows the work of Hal Varian, Google's chief economist, and Hyunyoung Choi, Google's senior economist, who suggest that nowcasting can be used to predict consumer behavior. Their paper, Varian and Choi (2009a), focuses on four research areas, one of which was "travel". The authors estimate the flow of travelers to Hong Kong from eight countries over a sample of several years. The authors used a simple autoregressive model, which incorporates data from Google search statistics. Varian and Choi showed that tourism in Hong Kong could be predicted accurately using the Google search statistics in conjunction with standard estimation measures. Their research has been validated elsewhere. For example, economists in Hungary, Istvan Janos Toth, and Miklós Hajdu (2012) followed Varian and Choi's nowcasting methodology, and created extremely refined estimates of consumer activity in that country. Their results show that Google can be a useful tool for nowcasting consumption in a country where Internet use is less than in developed countries.

In another paper, Varian and Choi (2009b) are interested in the relation between certain Google queries and filings for unemployment benefits. The "Initial Jobless Claims" is a weekly report issued by the US Department of Labor. It tracks the number of people filing for unemployment benefits and is considered a leading

indicator of the labor market. Using two labor market related categories in Google Trends from 2004 and initial unemployment benefit claims, they compare the trends over time and find that the time series show similar cyclical patterns trends since 2008. When adding the Google Trends series to the standard ARIMA model they found a positive correlation between searches related to “Jobs” and “Welfare & Unemployment,” which show the potential this mechanism has for improving prediction. Finally, when adding the Google Trends series, the model fit is improved significantly and the out-of-sample mean-absolute-error estimate decreased.

Askatas and Zimmerman (2009) and Suhooy (2009) examined similar unemployment data for Germany and Israel, respectively, and also found significant improvements in forecasting accuracy by using Google Trends.

Jeremy Gingsberg et al. (2009) focused web search queries data on the area of health. They created a method for analyzing large numbers of Google search queries to track influenza-like illness in a population. This research is particularly important because up-to-date influenza estimates may enable public health officials and health professionals to better respond to seasonal epidemics, which can lead to saving lives. The authors found that they can accurately estimate the current level of weekly influenza activity in each region of the United States, with a reporting lag of about one day. Additionally, they were able to consistently estimate the current influenza like illness (ILI) percentage 1-2 weeks ahead of the publication of official reports.

The term “nowcasting” is a contraction of “now” and “forecasting”. Nowcasting is a very helpful tool; it is useful beyond influenza epidemics. For example, the European Central Bank published a paper in December 2010 on how it uses nowcasting of large amounts of economic data to get an early jump on GDP estimates, which are of critical importance when orienting fiscal policy. Bánbura et al. (2010) stated that *“Nowcasting is particularly relevant for those key macroeconomic variables which are collected at low frequency, typically on a quarterly basis, and released with a substantial lag.”*

Most aggregate economic variables are published with a lag, and they are subsequently revised, making it difficult for policy-makers to make an accurate assessment of the current state of the economy. This issue makes nowcasting or predicting the present, extremely helpful for getting more timely forecasts on economic indicators. By using a nowcasting model we can achieve real-time data. Empirical evidence suggests that nowcasting is becoming a valuable tool for anyone interested in present economic and business conditions.

In conducting an empirical study for Chile, Carrière-Swallow and Labbé (2010) found problems related to apparent noise in the data of Google. These authors found that Google uses a sampling procedure to collect data, which implies that the same query performed on different days produce slightly different results. Using a panel of samples taken from Google search statistics, the authors eliminate noise in

variables and use the criterion of Rose, which states that a signal is readily detectable above its background noise if it reaches a level of five times the noise, coming to the conclusion that the underlying signal in the data is strong indeed. The authors use a basic model for predicting economic activity in the Chilean automotive sector. Despite the relatively low rates of Internet use in the country, they found that the addition of a Google search statistics variable improves nowcasting models and reduces the delay in the delivery of information.

Della Penna and Huang (2009) conduct a similar study to estimate aggregate consumption in the United States. They use information from Google's search statistics to construct an index of consumer confidence nationwide, and compare this index with existing indices. These authors use the ability to split the Google search data into various consumer categories, consisting of the final Google search-based index (SBI) comprising four components, and report that SBI is highly correlated with the Index of Consumer Sentiment (ICS) from the University of Michigan and the Consumer Confidence Index (CCI) from the Conference Board. They concluded that the SBI leads in time and predicts other indices. Additionally, in terms of forecasting consumer spending, the SBI outperforms the ICS and the CCI, thus providing independent information.

This research's objective is to examine whether Internet visits are related to actual visits and to demonstrate that Google searches are effective predictors of arrival of tourists to Punta Cana. Punta Cana may be a better case study than Hong Kong, used

in Varian and Choi (2009a), since Hong Kong is much more than just a tourist destination and people tend to search the Internet for the word “Hong Kong” for many other reasons, not necessarily to travel there. Furthermore, Punta Cana is known mainly as a touristic destination, and users are unlikely to perform a Google search of the word “Punta Cana” for any reason other than tourism. Since this case study is perhaps a better choice for testing this model, we’re hoping to get more accurate results than in the case of Hong Kong. Varian and Choi’s paper suggests that by using Google Trends data for a given region, highly accurate predictions can be made for such major economic activities like car and home purchases, and tourism decisions.

III. METHODOLOGY

3.1 Data

The Dominican Republic is an extremely popular vacation destination for travelers from all over the world. The country has several airports located in its most important communities. This analysis is based on the major tourist area of the country, defined as the area with the largest number of tourist arrivals and sustained economic growth in the period 2004-2013. The most attractive touristic city is Punta Cana, located on the east coast of the Dominican Republic. The fact that the Punta Cana International Airport (PUJ) receives the most tourists supports this decision. Punta Cana's importance for the Dominican Republic is so great that its international airport receives more flights than the country's capital, Santo Domingo. This is probably due to Punta Cana having an extensive network of luxury hotels, and many consider the beaches as one of the most beautiful of the Caribbean.

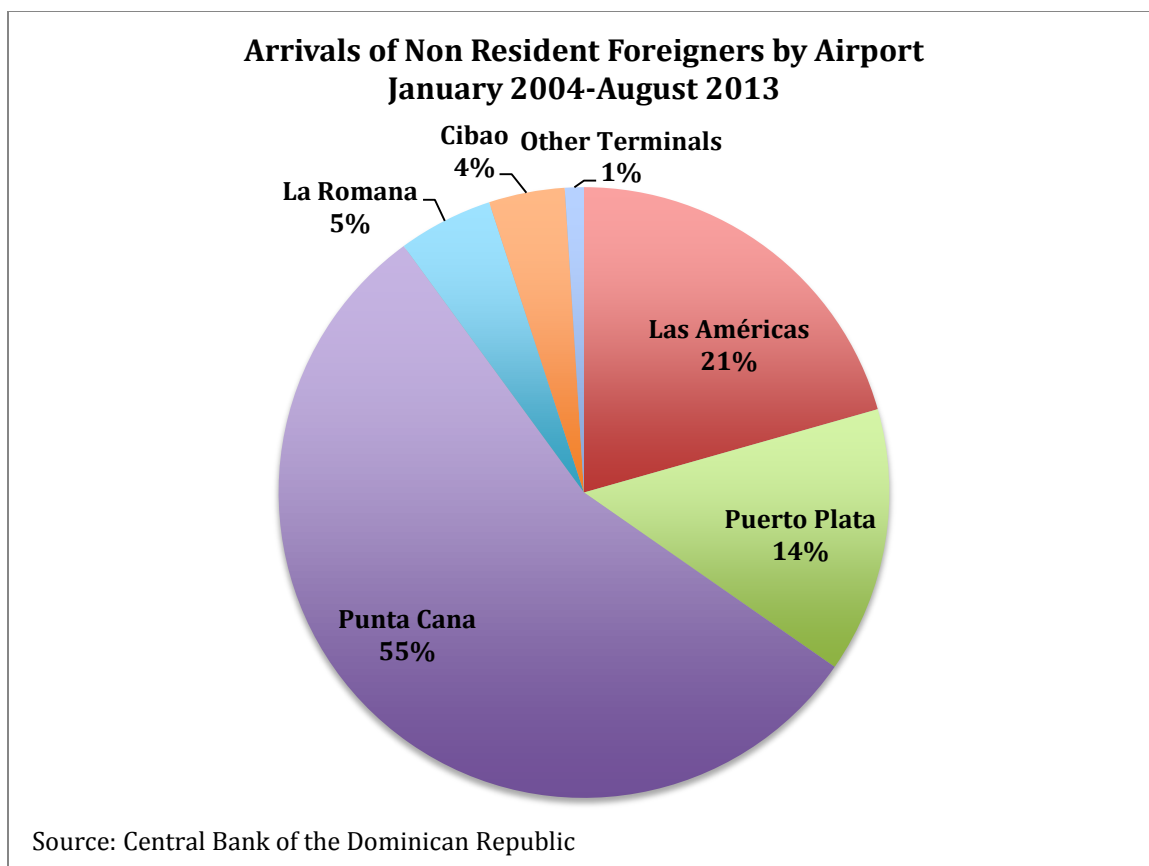


Figure 1: Arrivals of Non Resident Foreigners by Airport January 2004-August 2013

The statistics of arrivals of nonresidents foreigners, by airports, published by the Central Bank of the Dominican Republic, show that the Punta Cana International Airport has maintained supremacy in the reception of tourists to the country for the period in question. As shown in Figure 1, Punta Cana International Airport received the largest number of tourists. During the January 2004-August 2013 period, this airport received 18,553,098 visitors, equivalent to 55% of total tourist arrivals.

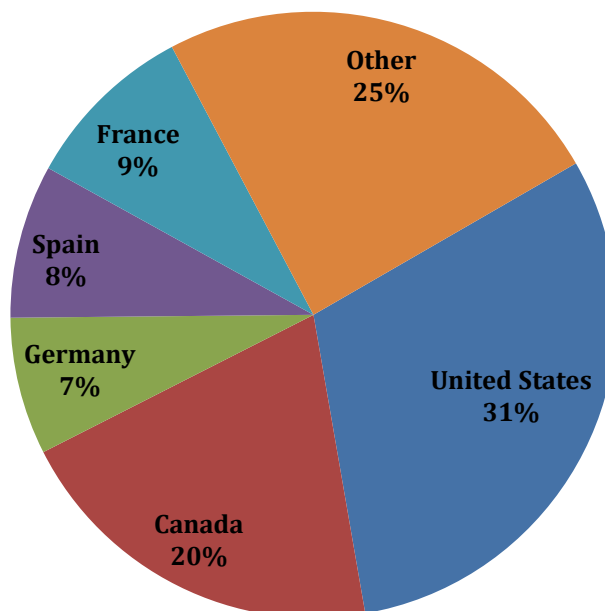
The Central Bank of the Dominican Republic records monthly arrivals to different airports by country of origin and passenger's condition since 1993. These data are

highly reliable, which are extracted through a daily count of the visitor cards of embarking or disembarking arrivals in the Dominican Republic.

During the January 2004-August 2013 period, the average arrivals of non-resident tourists from North America, accounted for 55.01% of all travelers. These arrivals are mainly from the United States and Canada. Mexico accounts for less than 0.5% of the North American total arrivals to Punta Cana.

Figure 2 depicts the composition of tourist arrivals to Punta Cana International Airport by country of origin; the United States and Canada are the top two countries from which most tourists depart to visit Punta Cana, accounting for 31% and 20% of total arrivals, respectively. Therefore, these two countries have been chosen for this analysis, being the leading countries in tourist arrivals.

**Composition of Punta Cana's visitors by country
Based on January 2004-August 2013 average, in %**



Source: Central Bank of the Dominican Republic

Figure 2: Composition of Punta Cana's visitors by country based on January 2004-August 2013 average, in %

3.2 The Google Index

Time series data on Google searches are available at weekly frequency from January 2004 onwards (<http://www.google.com/trends/>). The data provided by Google does not represent the number of searches in a given time period. Instead, they are presented as an index that compares the number of queries during a specified period with the global search activity of a given region. The data are normalized, with a base 0 January 1, 2004, in which the resulting numbers represent the percentage deviation from 2004, with a maximum value of 100 (Varian and Choi,

2009). In general, this recentering of the data should not affect the results, but it will be important for interpretation. It should be noted that Google Trends is sampled data, and changes slightly from day to day.

Google Trends analyzes a subset of Google web searches to compute how many of them have been done for the terms one enters, relative to the total number of searches done on Google over time. This analysis indicates the likelihood of a random user to search for a particular term from a specific location at a particular time. Keep in mind that Trends defines a threshold of traffic for search terms, so that those with low volume will not appear. The system also eliminates repeated queries from a single user over a short period of time, so that these types of queries do not artificially impact the variables of interest.

Google Trends also defines different categories. A “category” refers to a classification of industries or markets. When one selects a particular category, the data for the search term will be restricted to that category. For example, if you select the Travel category for the search term “cruise”, the data you see will be restricted to that specific category. If a user does not choose a category, then the data would be comprised across all categories. When one filters by category, the system only contains queries that are related to that category. In this research paper, we filter by the “Travel” category. Figures A3 and A4 show the Google Trends Interface for “Punta Cana” search activity in Canada and the United States, respectively.

3.3 Internet Usage

According to the World Bank, 86.8% of people in Canada have Internet access and 81.0% in the United States in 2012, where Internet users are people with access to the worldwide web. In addition, both Canada and the United States are developed countries. As shown in Figure 3, the ratio of Internet users per capita in these countries is much higher than the world average of 35.6%, which makes these countries good for this analysis.

When planning a vacation people tend to use the Internet to search for vacation/tourism related topics, like countries to visit, places to go, comparing and booking hotels, and so on. Therefore, it is safe to assume that the Internet is an important and well-used tool for vacation planning.

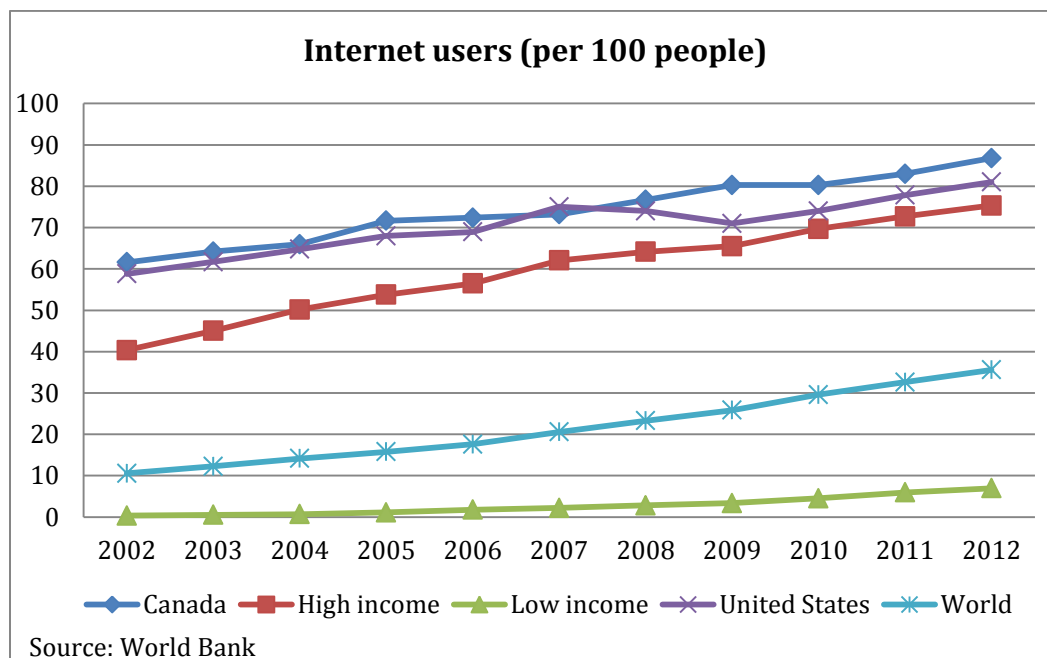


Figure 3: Internet users (per 100 people)

It should be noted that the reason Google is being used for this research is because it is the only search engine that provides these data for public use. There is no reason to dismiss the idea that the conclusions might concur for other web search engines, but this requires further research and additional data which are not available.

IV. MODEL SELECTION

To answer the question of whether Google is a good predictor of tourists visiting Punta Cana, we use the same model proposed by Hal Varian and Hyunyoung Choi for predicting the visits to Hong Kong. In that sense, we will have two regressions, one without the Google search index (the baseline model) and one with the search index as an additional predictor. We estimate the model using the method of ordinary least squares (OLS).

Before making any estimation we verify the characteristics of the data in terms of seasonality, which means that the data of tourists arriving to Punta Cana have a strong cyclical component, which is not exceptional for these types of variables.

We will employ a simple prediction model with logarithmic transformations, as Varian and Choi used, which is a seasonal autoregressive model that features values lagged at $t-1$ and $t-12$ (these are past values of the series itself, the month before and the year before, respectively) to predict current levels of arrivals. As we mentioned before, we will be using only two countries of origin for the analysis, which are the United States and Canada. We will first test if the current tourist arrivals to Punta Cana from the United States and Canada can be modeled using last month's tourist arrivals and last year's tourist arrivals. Then, we will add the Google trends variable that will represent the volume of queries made in a specific week of the month. The

week was chosen by running all weeks and choosing that of highest statistical significance in the model, and high correlation with the actual arrivals. Therefore, $(g_t can)$ represents the trend in the volume of queries done during the first week of each month during the January 2004-August 2013 period. Likewise, the same goes for $(g_t us)$ but by using the third week of each month for the Google Trends variable.

We will have a model for each country, wherein current tourist arrivals from the United States to Punta Cana will be denominated $(y_t us)$ and the current arrivals from Canada to Punta Cana will be denominated $(y_t can)$.

$$\log(y_t us) = \beta_1 + \beta_2 \log(y_{t-1} us) + \beta_3 \log(y_{t-12} us) + \varepsilon_t \quad (4.1)$$

Model (4.1) is the baseline model. Then we add the Google search variable $(g_t us)$, which is the standard index for the search volume of Punta Cana related queries over the 2004-2013 periods in the United States, obtaining (4.2):

$$\log(y_t us) = \beta_1 + \beta_2 \log(y_{t-1} us) + \beta_3 \log(y_{t-12} us) + \beta_4 \log(g_t us) + \varepsilon_t \quad (4.2)$$

Similarly, the models for Canada are:

$$\log(y_t can) = \beta_1 + \beta_2 \log(y_{t-1} can) + \beta_3 \log(y_{t-12} can) + \varepsilon_t \quad (4.3)$$

$$\log(y_t can) = \beta_1 + \beta_2 \log(y_{t-1} can) + \beta_3 \log(y_{t-12} can) + \beta_4 \log(g_t can) + \varepsilon_t \quad (4.4)$$

After estimating we will focus on showing that the model exceeds all typical problems that may face an OLS regression. We then compare the predictions from model (4.1) to model (4.2), and model (4.3) to model (4.4), and discuss the differences attributable to adding the Google variable. As explained before, we use a seasonal autoregressive model. This decision has strong intuition in the real world, as tourists usually travel at the same rate each year. After obtaining the estimation results and verifying compliance with the OLS assumptions, we would have the ability to predict the volume of tourists who can arrive to Punta Cana from Canada and the United States in the present period. Similarly, we would now have reliable dataset before the Central Bank of the Dominican Republic publishes such information. However, the purpose of this paper is to test if the Google variable actually helps in prediction, and the best way to do this would be to compare a model that does not include this variable with another model that does.

Hypothesis: This paper explores whether Punta Cana-related queries can be predictive of or correlated with the actual level of tourists that arrive at Punta Cana.

H₀: Including the Google Trends variable does not improve the predictive power of the baseline model.

H₁: Including the Google Trends variable does improve the predictive power of the baseline model.

All signs on the coefficients are expected to be positive. That is, visits to Punta Cana should be positively correlated with visits last month, last year and with the volume of searches on Google by tourists for both countries. Also, as the visitors' data has a lot of seasonality, the coefficient of visits in $t-12$ should be larger than the coefficient of visits in $t-1$.

V. ESTIMATION OF THE MODELS

Looking at the behavior of the data (Figures 4 and 5 in the Appendix), you can see that it shows seasonality; on average the data show maximum values every January for both countries and minimum values in June and September for Canada and the United States, respectively. The graphs might indicate that the period when there is the greatest travel-related interest in Punta Cana coincides with January, one of the months that have the coldest weather in these countries. After estimating both models we reached the following conclusions.

a. Canada

The estimated baseline model for Canada is shown in equation (5.1), for the period January 2004 to August 2013. The Google Trends variable (ca_g) for Canada was then added to the model and the estimation is shown in equation (5.2). We use the Google searches from the first full week (Sunday-Saturday) contained in the month. After testing weeks 1-3, we found that week 1 serves as best predictor, and shows a high correlation with the actual tourist arrivals in Punta Cana. The results are shown in Table A1 in the Appendix.

$$\log(y_t can) = 0.358 + 0.051 \cdot \log(y_{t-1} can) + 0.921 \cdot \log(y_{t-12} can) + \varepsilon_t \quad (5.1)$$

$$\log(y_t can) = 0.734 + 0.057 \cdot \log(y_{t-1} can) + 0.868 \cdot \log(y_{t-12} can) + 0.003 \cdot (ca_g) + \varepsilon_t \quad (5.2)$$

Models (5.1) and (5.2) return positive coefficients on all explanatory variables, which implies that Canadian tourist arrivals in $(t-1)$ and $(t-12)$ are positively related to Canadian tourist arrivals in t .

For the Canadian case we obtain that the Google trends variable is found to be statistically different from zero at the 10% and, the Adjusted R-squared has increased from one model to another (0.9724 to 0.9739). Model (5.1) indicates that variations in the independent variables explain 97.24% of the variance in the dependent variable. Model (5.2) indicates that 97.39% of the variations in the dependent variable is explained. We can also notice that all of the coefficient signs are positive, as expected. Therefore, the model that includes the Google trends variable has a better fit than the baseline model. We also test for the Akaike Information Criterion (AIC), used for a relative quality statistical model. This criterion is used as a means of model selection. As shown in Table A1 in the Appendix, the AIC criterion is smaller for Mod 2, which suggests that this model has a better goodness of fit. For the model in equation (5.2) we find that:

- A 100% increase in the arrivals of tourists in the previous month corresponds to a 5.7% increase in tourist arrivals to Punta Cana in this month.
- A 100% increase in the arrival of tourists in the previous year corresponds to a 86.77% increase in tourist arrivals to Punta Cana in this month.
- A 1 unit increase in the Google search volume is correlated with a 0.26% increase in tourist arrivals to Punta Cana this month.

b. United States of America

Similarly, we estimate the baseline model for the United States, equation (4.3) for the period January 2004 to August 2013. And then we estimate equation (4.4) using the logarithmic lagged variables, $\log(y_{t-1}us)$ and $\log(y_{t-12}us)$, adding the Google Trends index as a predictor. For the United States case, Google searches made in the third week of the month serves as best predictor. The results are shown in Table A2.

$$\log(y_tus) = 0.428 + 0.154 \cdot \log(y_{t-1}us) + 0.815 \cdot \log(y_{t-12}us) + \varepsilon_t \quad (5.3)$$

$$\log(y_tus) = 0.820 + 0.143 \cdot \log(y_{t-1}us) + 0.775 \cdot \log(y_{t-12}us) + 0.002 \cdot (us_g) + \varepsilon_t \quad (5.4)$$

We find that including the Google search statistics variable improves the nowcasting model for the United States. The model now explains 90.93% of the actual tourist arrivals, in comparison with the 90.76% without the Google index variable as a predictor. The Akaike Information Criterion (AIC) is again lower for the model that includes the Google Trends index. Again, we conclude that the model with the search statistics variable is a better fit.

When comparing these results with those obtained by Varian and Choi, specifically in the case of the variable of interest, (ca_g) and (us_g) the authors found a coefficient of 0.001 for the case of Hong Kong, which is smaller than the 0.002 coefficients found for the Punta Cana case. As expected, the Punta Cana case has relatively more precise coefficients than Hong Kong. As we can see in Table A1, the three criteria

used to conclude that the models including the Google index are more accurate for both the United States and Canada are met. The models present higher Adjusted R^2 , lower residual standard errors and smaller AICs for both countries.

VI. CONCLUSIONS

The main conclusion for this research is that tourism in Punta Cana, Dominican Republic can be accurately predicted using Google search statistics in conjunction with standard estimation measures. The results give better fits for both Canada and the United States.

Given the good fit of the models and the significance of the independent variables we can conclude that both models that include the Google index give more accurate predictions of the number of tourists arrivals to Punta Cana than the models that do not. We find that the Google trends variables for both the United States and Canada are statistically significant, but small in absolute value.

This research concludes that including the searches of “Punta Cana” that Canadians make on Google in the first week of a month helps predict the actual arrival of

Canadian tourists to Punta Cana in that month. The same applies in the case of Americans, but using the searches made during the third week of a month.

The improvement attributable to the inclusion of the Google trends variable was tested and supported by three criteria: an increase in the Adjusted R-squared, a smaller Residual Standard Error and a smaller Akaike Information Criterion (AIC). By analyzing the Adjusted R-squared we can see that it has increased from one model to another in both cases, which implies that variation in the dependent variable is better explained by the independent variables. Additionally, the residual standard errors decreased for both countries when the Google variable is included. These conclusions indicate that the models including the relevant Google Trend index, outperforms the models that exclude that variable.

Finally, it is considered of great importance for the study of empirical econometrics to use variables such as Internet searches, since it has been shown to help improve predictions in many areas of the economy.

VII. REFERENCES

Askatas, N. and K.F. Zimmermann (2009). Google econometrics and unemployment forecasting. IZA Discussion Papers 4201, Institute for the Study of Labor (IZA).

Carrière-Swallow, Y. and F. Labbé (2010). Nowcasting with Google trends in an emerging market. Working Papers of the Central Bank of Chile, N° 588.

Choi, H. and H. Varian (2009a). Predicting the present with Google trends. Technical report, Google Inc.

Choi, H. and H. Varian (2009b). Predicting initial claims for unemployment benefits. Technical report, Google Inc.

Della Penna, N. and H. Huang. Constructing consumer Sentiment Index for U.S. Using Internet Search Patterns (2009). Working Paper N° 2009-26, University of Alberta.

Gingsberg, J., M. H. Mohebbi, R. S. Patel, L. Brammer, M.S. Smolinski, and L. Brilliant (2009). Detecting influenza epidemics using search engine query data. *Nature* 457, 1012-1014.

Goel, S., J. M. Hofman, S. Lahaie, D. M. Pennock, and D. J. Watts (2010). Predicting consumer behavior with Web search. Technical report, Proceedings of the National Academy of Sciences.

Janos Toth, I. and Miklós Hajdu (2012). Google as a tool for nowcasting household consumption: estimations on Hungarian data. Institute for Economics and Enterprise Research- HCCI.

Meese, Richard and Kenneth Rogoff (1983). Empirical Exchange Rate Models of the Seventies: Do they fit out of sample? *Journal of International Economics*, 14: 3-24.

Suhoy, T. (2009). Query indices and a 2008 downturn: Israeli data. Technical report, Bank of Israel.

Tourist Inflow Report, Central Bank of the Dominican Republic, January-October 2013

Working with Google Trends/Understanding Features, Google Trends

APPENDIX

VIII. APPENDIX

Figure A1: Canada Arrivals and Google Trends

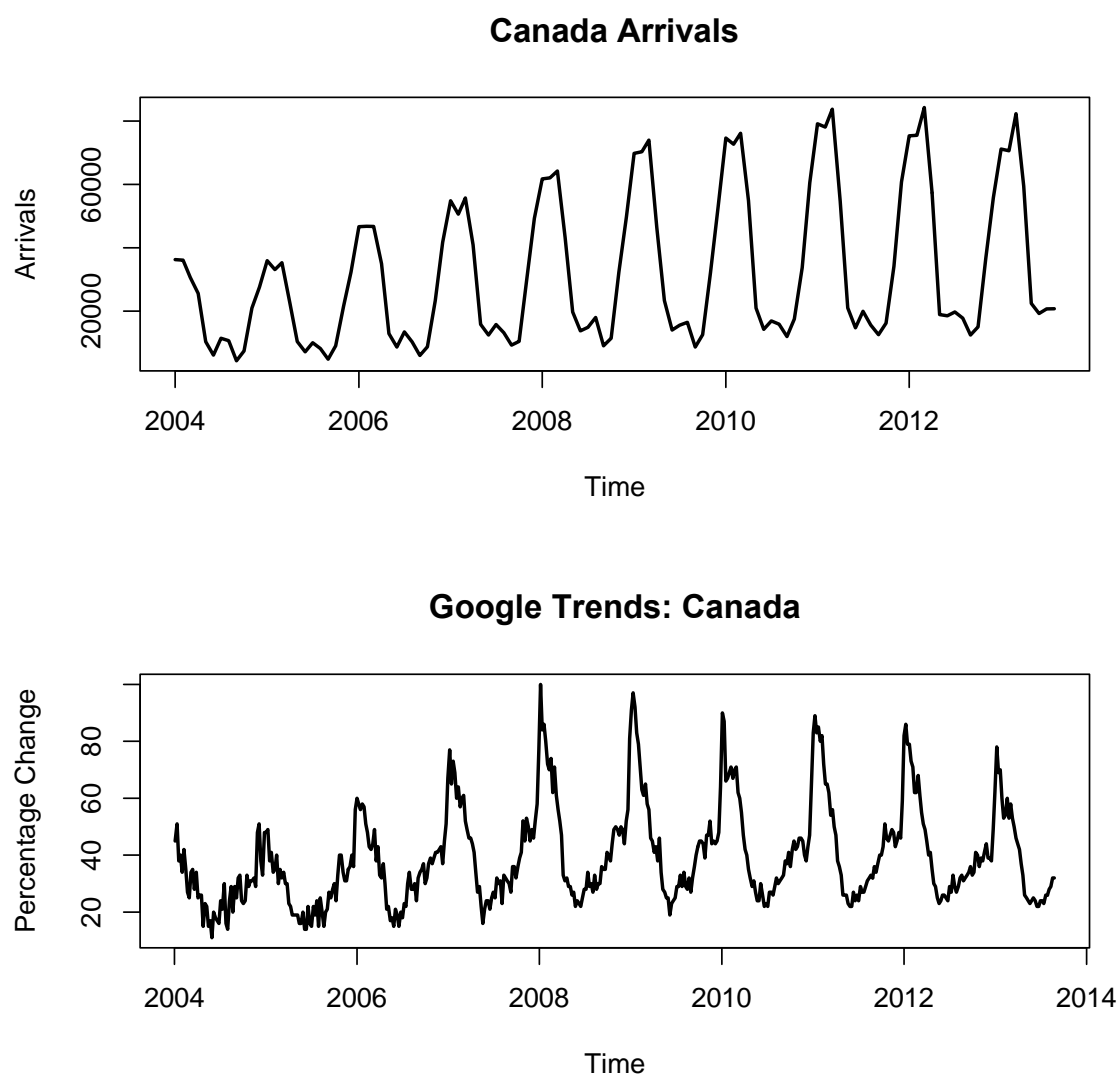


Figure A2: United States Arrivals and Google Trends

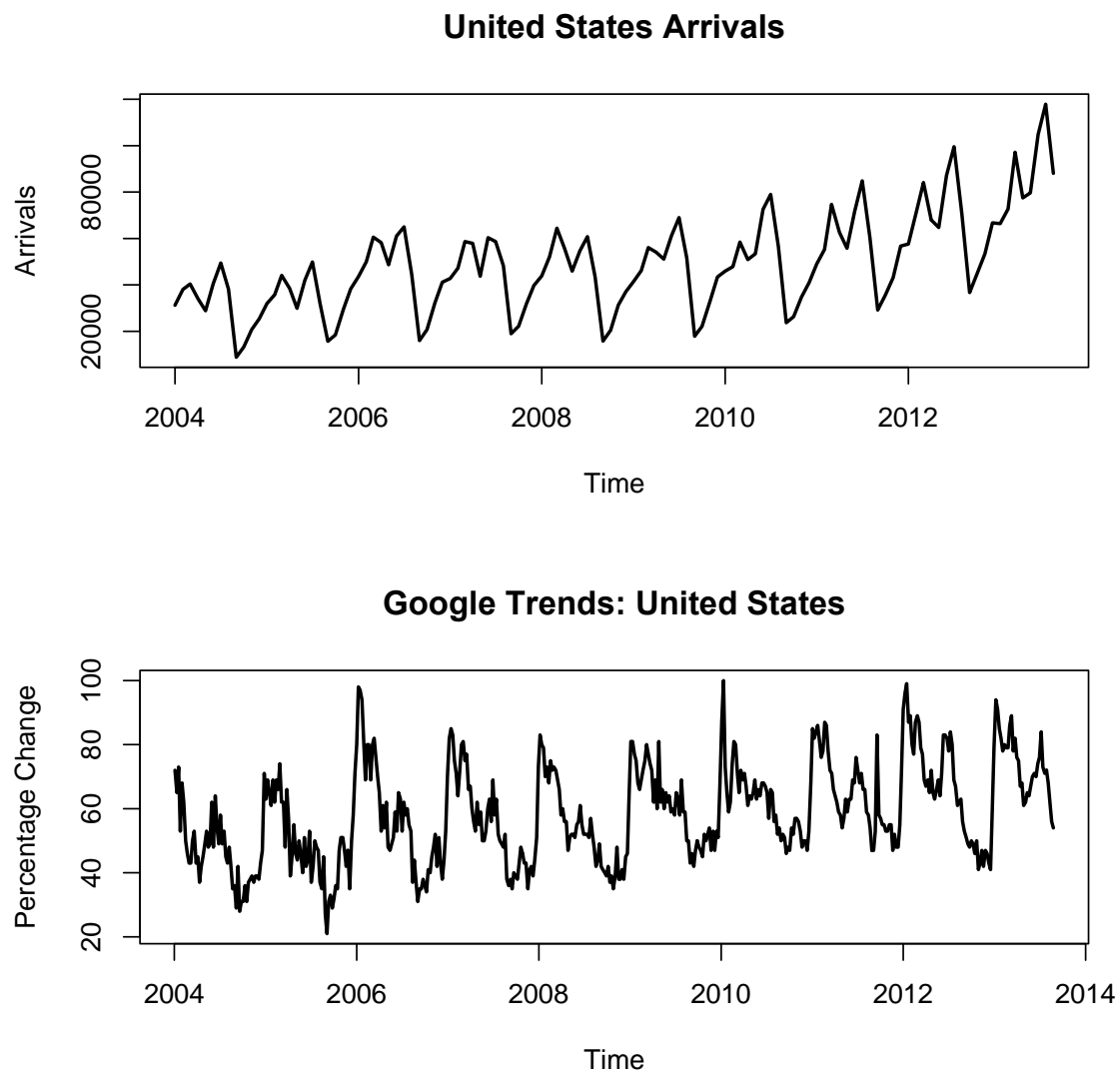


Table A1: Canada and United States Estimation Results

OLS-Nowcasting

2004.1 - 2013.8

Canada			United States		
Variable	Mod 1 logca	Mod 2	Variable	Mod 1 logus	Mod 2
C	0.358 * (0.174)	0.734** 0.225	C	0.427 (0.346)	0.82 * (0.384)
logca(-1)	0.051 . (0.027)	0.057* 0.026	logus(-1)	0.154 *** (0.04)	0.144 *** (0.04)
logca(-12)	0.921 *** (0.026)	0.868*** 0.033	logus(-12)	0.815*** (0.039)	0.775 *** (0.042)
cag week 1		0.003* 0.001	usg week 3		0.002 * (0.001)
Obs	103	103	Obs	103	103
Adj R ²	0.972	0.974	Adj R ²	0.9076	0.9093
RSE	0.124	0.121	RSE	0.132	0.13
AIC Info. Criteria	-133.05	-135.54	AIC Info. Criteria	-120.99	-121.85

Note: Values in parentheses indicate the standard errors. Signif. codes: 0 '***' 0.001
 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table A2: Canada and United States Descriptive Statistics
2004.01-2013.08

Descriptive Statistics				
Statistics	Arrivals		Google trends	
	US	CAN	US	CAN
Average	49,103	32,499	58	39
Median	47,534	22,041	57	35
Min	8,827	4,337	21	11
Max	117,868	84,301	100	100
Standar Desviation	20,564	23,106	15	17
Observations	116	116	504	504

Figure A3: Search index for “Punta Cana” in Canada

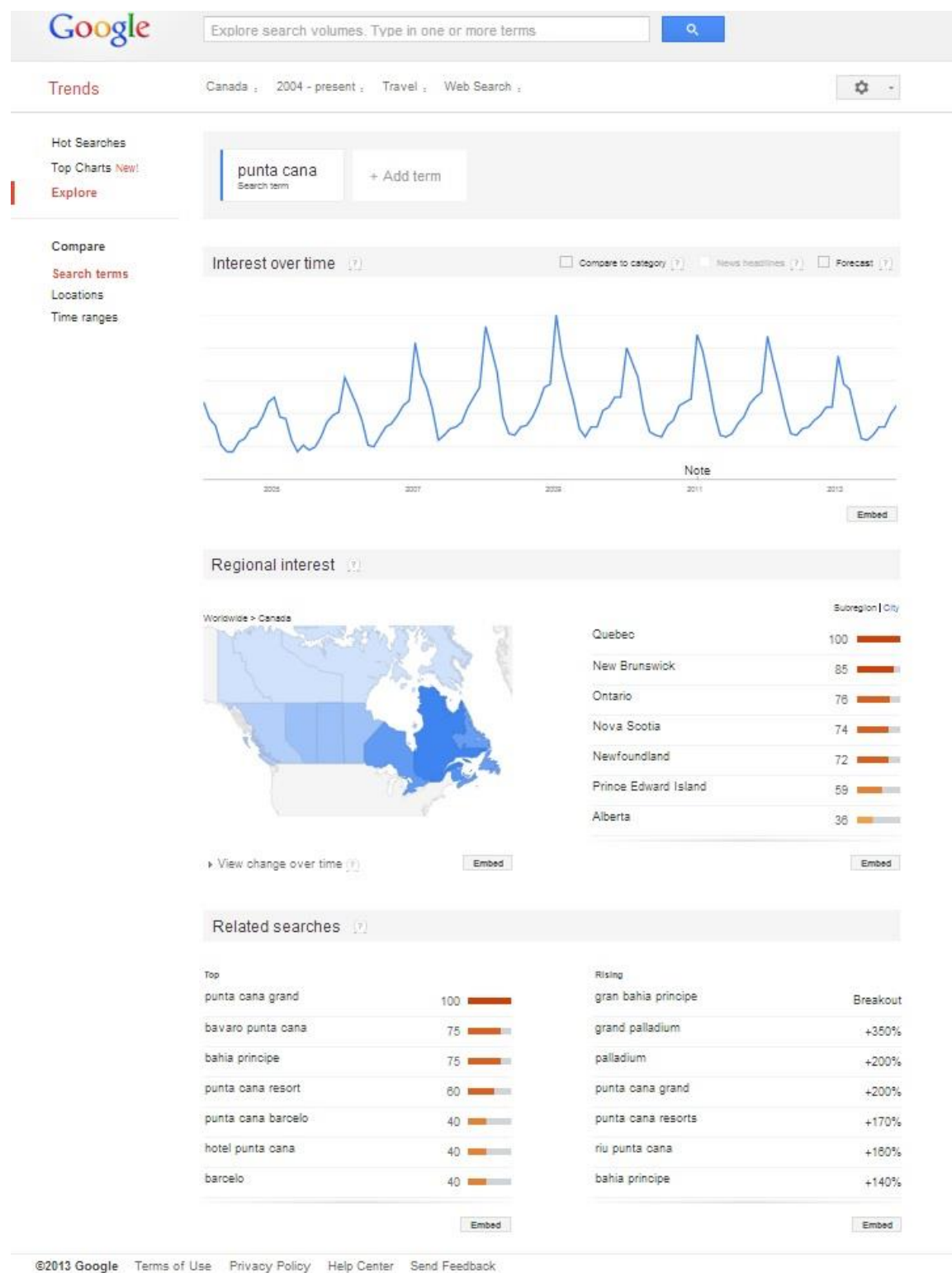


Figure A4: Search index for “Punta Cana” in the United States

